Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-02 23:28:50
REFERENCE NO: 203

Author Names & Affiliations

- Malin Pinsky - Rutgers University

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Biology, Ecology & Evolution

Title of Submission

The democratization of high performance computing

Abstract (maximum ~200 words).

High performance computing (HPC) is quickly becoming a necessary tool in fields not traditionally associated with "advanced cyberinfrastructure." Fields in biology and geosciences, from global change to population genetics, now require storing, manipulating, analyzing, and visualizing datasets too large for desktop computers. Widespread training in the use of high performance computing and access to high performance computing resources are critical for continued advances in these fields. The need for these skills and resources, however, extend far beyond those relatively few research groups that specialize in using HPC. NSF has an opportunity to train the next generation of scientists and spur more rapid research progress across many fields by ensuring that HPC and data science training and access are widely available.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

A sampling of scientific challenges that would benefit from a wider range of research groups with training in and access to HPC resources:
1) Ecological and evolutionary feedbacks in the earth's climate systems, including global ecological and evolutionary models, and the analysis of global datasets for marine and terrestrial ecosystems.
2) Ecological and evolutionary impacts of global climate change. To what extent will shifting distributions, evolution, physiological adaptation, and extinction play a role in biotic responses to climate change over the coming century? How will these processes differ across species, habitats, and ecosystems? Computational needs include both simulations and the analysis of global datasets.
3) Genomics of non-model organisms. What are the genomic mechanisms that underlie the incredibly diversity of life on earth? Answering these questions requires linking modern genomic techniques to the incredible store of natural history understanding in many non-genomics research groups. The techniques to generate genomic data are now available for just about any organism, opening entirely new views into

ecology and evolution in the wild. The cost of these techniques has also fallen so far that they are accessible to nearly any lab. However, the ability to analyze and the access to the computational resources continues to put these techniques out of reach for many.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

The need for more widespread access to HPC has been recognized within NSF, including the XCEDE system, and particularly the Comet system for serving the "long tail" of research computing by making access more rapid for moderate-scale projects.

Going forward, these efforts can and should be expanded further, publicized more widely, and accompanied by more widespread training, particularly at the graduate level. Classes along the lines of "Software Carpentry" are becoming more popular, but remain too rare to reach the number of students that need these skills. Current training opportunities rarely cover HPC, and so these resources remain underutilized, even by researchers who would benefit from them.

In addition, further efforts to lower the bar to HPC access will ensure that it is used by more of the "long tail," including students. These include rapid allocation of startup accounts, and ease of request for larger research or educational accounts.

Consent Statement

- “I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode).”